MgtOp 591 **Final Exam** **Summer 2008**

Name: _____

Instructions

- This test is a take home test and due 4 p.m., Friday, June 13, 2008. **No** late exam will be accepted under any circumstance.
- You should not receive any help to take this test or give any help to anyone taking this test. If you abide by this, at the end of your exam you should print the following statement and sign: "*To take this test, I have not received any help from anyone nor given any help to anyone taking this test.*" You forfeit your final exam without this statement on your test.
- Show **all** your work **clearly**.
- Each question is worth 5 points.
- If you use MS Excel/ PHStat or statistical programs such as Minitab or SAS for this exam, you are required to attach the necessary output. The output must be in the same order as the questions and properly indexed and referred to.

GOOD LUCK
AND
HAVE A WONDERFUL SUMMER

Name: _____

Problem 1:

You currently own a house in Pullman, Washington, and are planning on moving sometime in the next year. A friend of yours has suggested that you try to sell the house on your own. The decision is a difficult one. Real estate commissions can be quite hefty, but brokers bring a great deal of expertise to the process of selling a house, particularly the ability to set a reasonable asking price. Real estate professionals can quantify a variety of unobservable or unmeasurable characteristics of a house, such as that shiny purple wallpaper in the upstairs bathroom and the overall desirability of the neighborhood. To get an initial feel for the types of houses in your area, you have collected data that contain information such as the selling price of the house, square feet, numbers of bedrooms and bathrooms, and the presence of a fireplace for 108 homes sold in your neighborhood last year. The Excel file HOUSE contains the following variables. PRICE (y) is the selling price of the house in thousands of dollars SQ_FT (x_1) is the variable measuring the total square feet in the house. BEDS (x_2) and BATHS (x_3) are number of bedrooms and bathrooms, respectively. HEAT (x_4) and STYLE (x_5) are categorical variables. HEAT takes on the value of zero for gas forced air heating and one for electric heat. STYLE is the architectural style of the home: zero indicates a tri-level, one indicates a two-story house and two indicates that the house is a ranch-styled home. GARAGE (x_6) is the number of cars that can fit into the garage. We don't know whether it is an attached garage. AGE (x_7) is the age of the home in years. FIRE (x_8) and BASEMENT (x_9) indicate the presence (one) or absence (zero) of at least one fireplace or basement. SCHOOL (x_{10}) is the school district (0=South school district; 1=North school district). All else the same, the South school district is viewed as the preferred school district.

a. If you employ the forward selection method to select predictor variables, which variable among x_1 through x_{10} is the first one to be selected and why?

b. Suppose you fit a regression model using PRICE (y) as response variable and the rest of the variables described above as predictor variables. Write down the estimated regression line and find the estimate of the variance of the error term.

c. If you employ the backward elimination method to select predictor variables, which variable is the first one to be eliminated from the model in Question 1-b and why?

d. Is the regression model fitted in Question 1-b significant? Justify your answer? (You need to state the null and alternative hypothesis explicitly and draw your conclusion about the hypothesis.)

e. If the fitted model is significant, identity the estimated coefficients that are significant using the “experiment-wise” Type-I error of 0.1.

f. Is there a collinearity problem? Justify. If there is collinearity, identify predictor variable(s) that should be flagged.

g. It is known that the quadratic term of the size of the property is a useful predictor variable. Is there any indication that we may need the SQ_FT squared term, that is, x_1^2 in the model? Justify.

h. Is the homoscedasticity assumption about the error term appropriate in the model fitted for Question 1-b? Justify.

i. Is the normality assumption about the error term appropriate in the model fitted for Question 1-b? Justify.

j. If there are outlying observations, identify those? Justify your answer.

k. If you select predictor variables using the step-wise procedure, what is the fitted regression model?

l. You may have noticed that although the STYLE (x_5) is a categorical variable with 3 categories, it is used as if it were a numerical variable taking values 0, 1 , and 2. A more appropriate approach is to define indicator (dummy) variable(s) for the STYLE. Define indicator variables for the STYLE, and discuss under what situation you may use the STYLE as it is, that is, taking values 0, 1, and 2 instead of the indicator variable(s).

m. Glancing through the fitted model in Question 1-b, BATHS (x_3), STYLE (x_5), FIRE (x_8), and SCHOOL (x_{10}) are apparently insignificant. Test the null hypothesis $\beta_3 = \beta_5 = \beta_8 = \beta_{10} = 0$.

n. What is the alternative hypothesis in Question 1-m?

o. If the null hypothesis in Question 1-m is rejected, identify significant coefficients.

p. Identify influential observations, if any, based on the leverages (`HI1`) and Cook's distances (`COOK1`) on the next page obtained from Minitab.

Name: _____

Leverages and Cook's distances

n	HI1	COOK1	n	HI1	COOK1
1	0.166336	0.001209	55	0.058452	0.000004
2	0.297894	0.007848	56	0.070646	0.009103
3	0.099007	0.001254	57	0.028602	0.006993
4	0.061484	0.006122	58	0.050172	0.000849
5	0.152929	0.023668	59	0.167772	0.082502
6	0.078339	0.000482	60	0.045354	0.002506
7	0.156363	0.019668	61	0.111637	0.000009
8	0.139658	0.049961	62	0.034940	0.000002
9	0.125986	0.032759	63	0.075533	0.000570
10	0.082002	0.001804	64	0.182108	0.045665
11	0.123746	0.007941	65	0.068288	0.004325
12	0.143388	0.067431	66	0.115121	0.016464
13	0.139299	0.089279	67	0.054906	0.002354
14	0.282503	0.026979	68	0.068194	0.019490
15	0.189924	0.000264	69	0.065356	0.001393
16	0.103558	0.004918	70	0.041054	0.002907
17	0.067117	0.003697	71	0.179074	0.001677
18	0.105859	0.000102	72	0.121630	0.000639
19	0.115114	0.004133	73	0.046753	0.000390
20	0.140933	0.008059	74	0.063317	0.009441
21	0.308584	0.021037	75	0.029411	0.004695
22	0.077775	0.000087	76	0.070980	0.000296
23	0.090701	0.003958	77	0.037913	0.002880
24	0.098272	0.011495	78	0.065388	0.001525
25	0.069583	0.000046	79	0.063559	0.014927
26	0.130281	0.006316	80	0.071672	0.001980
27	0.090131	0.000269	81	0.041482	0.006948
28	0.102607	0.000117	82	0.056293	0.000727
29	0.085506	0.002916	83	0.062406	0.002668
30	0.064224	0.002020	84	0.040737	0.000011
31	0.091597	0.002445	85	0.190656	0.000010
32	0.070334	0.004246	86	0.037700	0.000024
33	0.091686	0.005311	87	0.085095	0.036601
34	0.107946	0.000267	88	0.029018	0.000600
35	0.095550	0.000756	89	0.075772	0.000010
36	0.096644	0.000374	90	0.082532	0.000406
37	0.161607	0.002028	91	0.044816	0.000492
38	0.091474	0.005824	92	0.100723	0.009539
39	0.175318	0.254289	93	0.170614	0.001176
40	0.101705	0.000527	94	0.296753	0.159896
41	0.138103	0.024125	95	0.089050	0.008400
42	0.181985	0.001982	96	0.072076	0.015921
43	0.088373	0.000368	97	0.111135	0.001510
44	0.147906	0.002302	98	0.103528	0.007962
45	0.209241	0.074631	99	0.106460	0.002686
46	0.115495	0.004638	100	0.086103	0.000705
47	0.053271	0.001873	101	0.076705	0.001022
48	0.155993	0.000644	102	0.091275	0.000665
49	0.094959	0.007465	103	0.121593	0.000243
50	0.063445	0.000001	104	0.039897	0.015817
51	0.057016	0.000532	105	0.068584	0.009875
52	0.082071	0.014498	106	0.053649	0.017665
53	0.057404	0.005051	107	0.062046	0.017430
54	0.042947	0.000202	108	0.154294	0.064407

Name: _____

Problem 2:

The data in the Excel file RedWine are the quarterly red wine sales in Australia during the period of first quarter of 1980 through the second quarter 1995.

a. Examine a time series plot of data and explain why a log transformation of the data may be needed.

b. Fit the following model and report the estimates of the model parameters and the estimate of the variance of the error term.

$$y_t = \beta_0 + \beta_1 t + \beta_2 Q_2 + \beta_3 Q_3 + \beta_4 Q_4 + \varepsilon_t,$$

where y_t is the (natural) log of the quarterly wins sales, t is the time variable, and Q_i for i=1, 2, 3 is the indicator variable such that $Q_i = 1$ for the i^{th} quarter and zero other wise.

c. Is the assumption of independence of the error term ε_t valid in the model of Question 4-b? Justify your answer using the Durbin-Watson test.

d. ~~If the independence assumption is violated in Question 2-c, fit a more appropriate model and provide the fitted model.~~

e. We let x_t be the fourth difference of y_t, that is, $x_t = y_t - y_{t-4}$ known as the quarterly difference. Then fit an ~~AR(4) model to this fourth differenced series~~ x_t, and provide the fitted model.

f. Based on the model you fit in Question ~~4-d~~ 2-b, find the forecasts of the red wine sales for the third and the fourth quarter of 1995.

Name: _____

Problem 3.

A bank branch located in a commerce district of a city has developed an improved process for serving customer during the 12 p.m. and 1 p.m. peak lunch period. The waiting time in minutes (operationally defined as the time the customer enters the line to when he or she is serve) of all customers during this hour is recorded over a period of one week. A random sample of 15 customers is selected, and the data are in the Excel file BANK1 in the CD.

a. Obtain the following summary statistics:
   i) Mean
   ii) Standard deviation
   iii) Median
   iv) Coefficient of variation
   v) Inter quartile range
   vi) Standard deviation of the sample mean
   vii) Mean absolute deviation

b. Are the waiting times from a normal population, i.e., a normal distribution? Justify your answer by
   i) stating the null and the alternative hypotheses;
   ii) reporting the test statistic;
   iii) either providing the critical value or the observed significance level; and
   iv) stating your decision.
   Use a significance level of 0.05.

c. The goal of the improved process was to make the average waiting time during the peak lunch period less than five minutes. Do the data provide strong evidence that the goal has been met? Justify your answer using a significance level of 0.05.

Problem 4

Suppose that another branch, located in a residential area, is most concerned with the Friday evening hours from 5 p.m. to 7 p.m. The waiting time in minutes of all customers during these hours is recorded over a period of one week and stored in the Excel file BANK2 in the CD.

a. Are the variances of the waiting times of the two branches, in the commercial district and in the residential area are the same? Justify your answer using a significance level of 0.05.

b. Are the means of the waiting times of the two branches, in the commercial district and in the residential area are the same? Justify using a significance level of 0.05. Your justification must be consistent with your answer in Question 4-a.

Name: _____

Problem 5

A marketing manager is trying to set strategy for a marketing campaign for a new product line consisting of three rowing machines: the basic model, the designer model, and the complete model. To write the information brochure and press release, the manager need to know which model each type of consumer will prefer so that, for example, a sales person does not go overboard how practical a model is when the market actually consists of impulsive types. Data were gathered in a small test market. For each purchase, two variables were recorded and summarized below. One is the model (basic, designer, or complete) and the other is the type of consumer (practical or impulse).

	Practical	Impulse
Basic	22	25
Designer	13	18
Complete	54	19

a. Is the consumer type independent of the model purchased? Justify at $\alpha = 0.05$.

b. Would the test of homogeneity be more appropriate? Explain.

Name: _____

Problem 6

HiTech Engineering, Inc. (HEI) designs and manufactures a variety of industrial products. HEI sells its products directly to existing customers and also uses authorized distributors around the globe to assist in sales and distribution. HEI gets word to the market about a new product in various ways. For instance, sometimes they purchase magazine advertisements and sometimes they issue news release. Because HEI produces industrial products and not consumer products, it never uses electronic media for the purpose of advertising. HEI management would like to understand how effective each of the five basic types of advertising they've used has been. The five basic types are magazines, postcard decks, editorials, news release, and literature. Requests for information by prospective customers are called "leads," and the management would like to measure the "effectiveness" in terms of "leads." An analyst collects data on "leads" for the five types of advertising and stores in the Excel file HiTech. The five basic types, magazines, postcard decks, editorials, news release, and literature are coded as 1, 2, 3, 4, and 5, respectively, in the file. The analyst uses the analysis of variance model

(1) $$Y_{ij} = \mu_j + \varepsilon_{ij}, \quad \text{for } j = 1,\ldots,5 \text{ and } i = 1,\ldots,17.$$

a. Define these variables, Y_{ij}, μ_j, and ε_{ij} properly within the context of this problem.

b. State the necessary assumption(s) for model (1).

c. Is the "effectiveness" measured by "leads" the same for all five types of advertising? Justify your answer. (You need to state the null and alternative hypothesis explicitly and draw a conclusion about the hypothesis.)

d. If the "effectiveness" is not the same for all types of advertising, which types are different in the mean leads? Justify your answer. If the "effectiveness" is the same for all types, indicate so and skip this part.

e. Are the assumption(s) you stated in Question 6-b satisfied?